Niveau(x) d'enseignement: 3 Numéro EDUQ : 8270382
Centre d'intérêt: 74

Genesee, Fred.

Evaluation of the 1974-75 grade 5 French immersion class / F. Genesee, S. Chaplin. -- [Montréal] : Commission des écoles protestantes du Grand Montréal, Curriculum Department, 1975.
34 f. ; 28 cm.

"A report submitted by the Curriculum Department to the Protestant School Board of Greater Montreal". Bibliographie: f. 22.

Cote Dewey: 372.654/G327e Cote LC: LB/1528

Descripteur(s): *Français (Langue) *Langue seconde *Classe d'immersion *Evaluation de l'apprentissage *Evaluation de programmes *Développement de la personnalité *Primaire (Niveau)

Présentation des résultats d'une étude réalisée dans le cadre de l'évaluation d'un programme d'immersion en langue française destiné à des élèves anglophones de cinquième année. Il s'agit de comparer les habiletés en français, en anglais et en mathématiques ainsi que le développement de la personnalité d'élèves inscrits dans des classes d'immersion en 1974-1975 avec la situation de groupes témoins des secteurs anglophone et francophone de l'enseignement.

Des élèves de huit écoles de la région de Montréal (PSBGM et CECM) ont participé à l'enquête: 46 élèves inscrits dans des classes d'immersion, 64 élèves anglophones suivant l'enseignement régulier, enfin 58 élèves du secteur francophone. Les tests suivants ont été administrés en mai et juin 1975 et, pour deux d'entre eux, en novembre 1974: trois sections du "California achievement test", un exercice de composition anglaise, les tests de rendement en français et en mathématiques de la CECM, des tests de production orale, de compréhension orale et de lecture, enfin le "Security test", spécialement conçu aux fins de cette étude.

Sans toutefois atteindre le niveau des élèves francophones, les élèves des classes d'immersion ont obtenu d'excellents résultats, supérieurs à ceux du groupe témoin anglophone, en ce qui concerne la maîtrise du français, cela en particulier au niveau de la communication interpersonnelle. Aucune différence significative n'a été découverte au niveau de la maîtrise de l'anglais entre le groupe témoin et les classes d'immersion. Les résultats au "Security test" indiquent que les élèves des classes d'immersion font preuve d'une confiance et d'une maturité supérieures à celles des élèves du groupe témoin. L'hypothèse voulant que le passage par une classe d'immersion soit la cause de problèmes psychologiques n'apparaît donc pas fondée.

Type de document: r
Région(s) administrative(s): 063

Financement: Bénéficiaire - BEPGM
Organisme(s) source(s) - MEQ - DGEES
Programme(s) de subvention - PARC
Auteur de l'analyse: ar

# Evaluation of the 1974-75 Grade 5 French Immersion Class.

F. Genesee
S. Chaplin

E D U Q 8270382

# Evaluation of the 1974-75 Grade 5 French Immersion Class

F. Genesee

S. Chaplin

A report submitted by the Curriculum Department
to
The Protestant School Board of Greater Montreal

October 1975

ACKNOWLEDGEMENTS

I would like to express my thanks to the principals, teachers and students in the participating schools for their cooperation and interest during the course of this research. In addition, I would like to thank La Commission des Ecoles Catholiques de Montréal for permission to do testing in some of their schools; in this regard a special note of thanks to the principal and director of studies and their staff in the French school which participated.

Finally, I would like to thank Mr. M. Stanley, Director of Curriculum, for his comments and suggestions on a draft of this report.

TABLE of CONTENTS


Page

INTRODUCTION . . . . . 1

THE SAMPLE . . . . . 3

TEST BATTERY:

English Language Tests . . . . . 6

French Language Tests . . . . . 8

RESULTS:

English Language . . . . . 11

French Language . . . . . 16

CONCLUSIONS AND SUMMARY . . . . . 20

REFERENCES . . . . . 22

APPENDICES:

(A) 1973-74 Grade 4 English language results. . . . . 23

(B) 1973-74 Grade 4 French language results. . . . . 24

(C) Explanation of Statistical Terminology. . . . . 25

(D) Explanation of Stanine Scale. . . . . 27

(E) Security Test. . . . . 28

The current investigation is the third in a series of investigations being carried out to assess the progress of a group of students in elementary school who are following a French immersion program (see Polich, 1973, for the results of the K to 3 evaluations; and Genesee, Morin & Allister, 1974, for the results of the grade 4 evaluation). Because the same students, more or less, have been the subjects of these investigations over the last few years, we will be able to paint a picture of these students' development over the entire course of their elementary school years (K to 6). This report will present the findings from grade 5.

We found in last year's evaluation, when the students were in grade 4, that the Immersion students performed as well as or better than, in some instances, the non-immersion students on all of the English language tests--reading, word knowledge, arithmetic problem solving and concepts, and writing. The one exception to this was on spelling, where the Immersion students were rated somewhat more poorly than the non-immersion students. On the battery of French language tests, the Immersion students scored significantly higher than the non-immersion students on reading, language usage (Test de Rendement en Français), listening comprehension, oral production and mathematics (Test de Rendement en Mathématiques). When compared to a control group of native-French speaking students, the Immersion students scored remarkably well. For example, on the two Tests de Rendement, the Immersion group scored in the average stanine range; their reading comprehension skills, listening comprehension skills and some of their oral production skills were rated equivalent to those of the French students.

It was also found in last year's evaluation that acquisition of interpersonal communication skills in French--listening comprehension and oral production--did not depend upon the students measured I.Q.

However, acquisition of academic language skills -- reading, language usage and mathematical terminology -- was influenced by I.Q. level.

The relationship between I.Q. and second language acquisition has not been investigated in other French immersion projects. Otherwise, the results from last year's grade 4 evaluation are consistent with those from the St. Lambert project (Lambert & Tucker, 1972), and with those from other centres where elementary French immersion is being offered; for example, Ottawa (Edwards & Casserly, 1973) and Toronto (Barik, Swain & McTavish, 1973). A summary of last year's results is presented in Appendices A and B. There were three areas of investigation in the current study: 1) development of English language skills; 2) the acquisition of French language skills, and 3) personality development and, in particular, the development of "security" (see Appendix E).

## THE SAMPLE

The sample consisted of three groups:

1) grade 5 anglophone students in the French immersion program (Immersion);
2) grade 5 anglophone students in the regular English program with French-as-a-second language courses (English Control);
3) grade 5 francophone students in a regular French program from a school under the jurisdiction of La Commission des Ecoles Catholiques de Montréal (French Control).

A list of the specific English schools from which students were selected is presented in Table 1. The anglophone students were subdivided into three I.Q. categories:

1) above average: full scale I.Q. of 116 or above;
2) average: full scale I.Q. between 96 and 115; and
3) below average: full scale I.Q. below 95.

Classification was based on the students' score on the Canadian Lorge-Thorndike Test of Intelligence (1967) which they had written in October 1973. No I.Q. information was available for the francophone students. However, we were assured by the principal of the school and by the students' teachers that this sample of students represented an "average" group. Their performance on the Tests de Rendement en Français and en Mathématiques suggests that they are, in fact, somewhat above average-- Français, 6.7 stanine and Mathématiques, 6.4 stanine.

A summary of the average I.Q. scores and sample sizes is presented in Table 1. Because the 1974-75 grade 5 immersion class was relatively small, it was not possible to control the number of male and female students in our samples. Furthermore, the 1974-75 grade 5 immersion class was somewhat select and, therefore, there were relatively few "below average" students.

In fact, there were so few male and female students in the "below average" Immersion group that we were not able to distinguish between males and females in the statistical analyses of the test results.

## TABLE I

(A)

| Immersion Schools | English Control Schools |
|---|---|
| Edinburgh | Edinburgh |
| Meadowbrook | Meadowbrook |
| Willingdon | Willingdon |

(B) SAMPLE SIZES

| I.Q. | Sex | Immersion | English Control | French Control |
|---|---|---|---|---|
| Above Average | Male<br>Female | 11<br>10 | 9<br>10 | |
| Average | Male<br>Female | 7<br>11 | 12<br>13 | 30<br>28 |
| Below Average | Male<br>Female | 3<br>4 | 10<br>10 | |

(C) AVERAGE I.Q. SCORES

| | IMMERSION | ENGLISH CONTROL | TOTAL |
|---|---|---|---|
| ABOVE AVERAGE | 124.3 | 120.3 | 122.3 |
| AVERAGE | 107.2 | 106.2 | 106.7 |
| BELOW AVERAGE | 94.4 | 91.9 | 93.2 |
| TOTAL | 108.6 | 106.2 | 107.4 |

## TEST BATTERY

Three types of test were administered: (1) English language tests; (2) French language tests; and (3) a Security test. Both the Immersion and English Control groups took all 3 types of tests, while the French Control group took only the French language tests. Most of the testing was carried out in May and June, 1975; two of the tests were administered in November, 1974.

### English Language Tests

(1) Writing - Each student was asked to write a composition based on one of three topic sentences. The choice of topic sentence was left entirely to the student. The students were given 45 minutes to write the composition. During the first 10 minutes, they were told to choose a topic and to write down all the ideas they had for a story. They were then given 30 minutes to write the composition. During the last 5 minutes they were encouraged to go back and check their work.

Each composition was later marked independently by two grade 5 teachers from the P.S.B.G.M. along 9 scales --- 1) spelling, 2) sentence accuracy, 3) sentence complexity & variety, 4) organization, 5) originality, 6) overall, 7) length, 8) vocabulary, and 9) punctuation. On the first 6 scales, the ratings ranged from 5, indicating excellent performance for a grade 5 student, to 1, indicating totally inadequate performance for a grade 5 student. A rating of 3 indicates a level of performance to be expected of a grade 5 student. The compositions were graded in a random order and the raters did not

know the school group or I.Q. category of each author. The ratings for both raters were averaged to form a single score for each student on each dimension.

There were only two response categories for the length and punctuation scales - - "appropriate" and "inappropriate". There were three categories for the vocabulary scale --- "superior", "average", and "below average". The average percentage frequency of occurrence of each rating was calculated for each of these dimensions.

(2) Reading - The Reading Comprehension subtest of the California Achievement Test (C.A.T.), Level 3 Form A (1970), was administered. This is a group test consisting of 2 reference skill problems and 4 prose passages, with multiple-choice questions. Raw scores were analyzed.

(3) Vocabulary - The Vocabulary subtest of the C.A.T., Level 3 Form A, was administered. This is a timed group test with multiple-choice questions which require the student to choose synonyms for target words. Raw scores were analyzed.

(4) Spelling - The Spelling subtest of the C.A.T., Level 3 Form A, was administered. This is a timed group test, again with a multiple-choice question format. The student is asked to determine which word, if any, of a group of 4 words is misspelled. Raw scores were analyzed. All three subtests of the C.A.T. were administered in one sitting, making the total testing session approximately one hour.

## French Language Tests

(1) Reading

A modified version of the Test de Lecture "California": 4è année (1967) was administered. This is a timed group test consisting of four subtests; (1) antonymes (2) synonymes (3) directives à suivre and (4) interprétation des textes. For purposes of data analysis the raw scores from subtest (1) and (2) were combined to form a single score on vocabulaire and raw scores from subtests (3) and (4) were combined to form a single score on compréhension. Maximum possible score on vocabulaire is 28, and on compréhension is 30. Raw scores were analyzed.

(2) Test de Rendement en Français (Niveau V)

This is a standardized group test devised by La Commission des Ecoles Catholiques de Montréal. It consists of 38 multiple-choice questions assessing the following kinds of language skills: vocabulary, grammar, spelling, etc. Norms, in the form of stanines, are established in November of each academic year when all students at level 5 in the C.E.C.M. write this test. In order that administration of this test to our students should coincide with administration of the test in the C.E.C.M., this test was given in November 1974. Stanine scores were used in the statistical analysis.

(3) Test de Rendement en Mathématiques (Niveau V)

This is also a standardized group test devised by the C.E.C.M.; again, norms are established in November of each year. Thus, this

test was administered in November 1974 to coincide with C.E.C.M. testing. It consists of 40 multiple-choice items. It must be remembered when interpreting the results of P.S.B.G.M. students on this test that it was devised with the mathematics curriculum of the C.E.C.M. in mind. In the present study, raw scores were converted to stanines for statistical analysis.

(4) Compréhension Aurale

This test was designed and created in the Curriculum Department of the P.S.B.G.M. It consisted of a short, 5-minute story entitled "La Petite Caroline" and 17 related multiple-choice questions. It is a group test. The story was presented first of all in its entirety to the students by means of a reel-to-reel tape recorder; it was then retold interjecting one or two questions after selected passages. Raw scores were analyzed.

(5) Production Orale

The nature of this test, individual, and the time required to score it, to be described below, precluded testing all students in the samples. Consequently, subsamples of students from each school, sex and I.Q. grouping were selected and interviewed (See Table 2 for a summary of the sample sizes). The test took the form of an interview between each select student and a female native-French speaker. Conversation was initiated by having the children describe a cartoon picture; following this, the interviewer engaged the child in a conversation about the cartoon and about matters of interest to the child. The interviewer was not told which I.Q. group each child belonged to. All interviews were tape recorded for later analysis.

TABLE 2

SAMPLE SIZES FOR
ORAL PRODUCTION EVALUATION
I.Q.

| | Above Average | Average | Below Average |
|---|---|---|---|
| Immersion | 13 | 9 | 6 |
| English Control | 11 | 13 | 11 |
| French Control | | 34 | |

Each student's speech sample was rated independently by two native-French speakers, using a 5-point rating scale, along 5 dimensions: compréhension aurale, prononciation, grammaire, vocabulaire and communication. The dimension of communication was designed to assess the children's willingness to speak French and their ease while doing so. A rating of 5 on a particular dimension indicates excellent or native-like performance on that dimension while a rating of 1 indicates very poor performance. The ratings of both raters for each student were averaged to give one score on each dimension. In last year's evaluations, we found that there was extremely high inter-rater reliability so that we did not feel it was necessary this year to analyse each rater's ratings separately. The raters did not know the I.Q. classification or school program of the students.

On occasion we are asked whether the French immersion program has any "psychological effects" on the students; in other words, how do students in French immersion react from the point of view of emotionality or personality? This type of question is very difficult to answer because of the lack of valid and suitable testing instruments related to these issues. Nevertheless, in an attempt to gather some information about the psychological effects of French immersion on the participating students, we administered a test designed to assess feelings of security to the grade 5 English samples. This test was designed especially for grade 5 students so that we felt that it was more adequate that other tests of its type. The results from this investigation are presented as Appendix E.

## RESULTS

A brief explanation of the statistical terminology used in the remainder of this report is presented as Appendix C.

### English Language Tests

Separate two-way analyses of variance (unweighted means solution for unequal n) were performed on the raw scores form the California Achievement Test-- vocabulary, reading and spelling. The main variables were School Group (Immersion, English Control) and I.Q. (above average, average and below average). A summary of these analyses is presented in Table 3.

#### Vocabulary

There was no significant difference between the Immersion ($\bar{X}$=32.96) and English Control ($\bar{X}$=31.24) students on this test. In terms of percentile, the average group performance was 68% (this score is based on an American student population).

As expected, there was a significant I.Q. difference with the above average students ($\bar{X}$=35.50) scoring higher than the average students ($\bar{X}$=31.84) who scored higher than the below average students ($\bar{X}$=28.95).

#### Reading

There was a significant difference between the groups on this test-- the Immersion students ($\bar{X}$=29.03) scored higher than the English Control students ($\bar{X}$=26.70). There was no such difference between these two groups on last year's English test. In terms of percentile, the Immersion group scored at the 57% level, and the English Control group scored at the 48% level.

TABLE 3

SUMMARY OF ENGLISH LANGUAGE TEST RESULTS

| | (A) by School Group | | | | (B) by I.Q. | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Immersion | Control | F Ratio | df | Above Average | Average | Below Average | F Ratio | df |
| **California Ach. Tests:** | | | | | | | | | |
| Vocabulary | 32.96 | 31.24 | 2.07 | 1,95 | 35.50 | 31.84 | 28.95 | 10.04*** | 2.95 |
| Reading | 29.03 | 26.70 | 4.12* | 1,95 | 32.19 | 26.96 | 24.45 | 15.84*** | 2,95 |
| Spelling | 23.03 | 22.70 | .09 | 1,95 | 25.36 | 23.17 | 20.06 | 7.76*** | 2,95 |
| **Writing Skills:** | | | | | | | | | |
| Spelling | 3.40 | 3.40 | .001 | 1,98 | 3.51 | 3.42 | 3.29 | .36 | 2,98 |
| Sentence Accuracy | 2.89 | 2.94 | .19 | 1,98 | 3.22 | 2.95 | 2.37 | 9.72*** | 2,98 |
| Sentence Complexity | 2.62 | 2.67 | .16 | 1,98 | 2.95 | 2.63 | 2.37 | 7.25*** | 2,98 |
| Organization | 2.81 | 2.77 | .05 | 1,98 | 3.05 | 2.86 | 2.46 | 6.61** | 2,98 |
| Originality | 2.10 | 2.08 | .05 | 1,98 | 2.28 | 2.11 | 1,89 | 5.03** | 2,98 |
| Overall | 2.82 | 2.83 | .001 | 1,98 | 3.17 | 2.79 | 2.51 | 7.10*** | 2,98 |
| **Punctuation:** Appropriate | 77% | 84% | | | 92% | 79% | 43% | | |
| Inappropriate | 23% | 16% | | | 8% | 21% | 57% | | |
| **Length:** Appropriate | 69% | 65% | | | 88% | 64% | 43% | | |
| Innappropriate | 31% | 35% | | | 12% | 36% | 57% | | |
| **Vocabulary:** Superior | 11% | 9% | | | 18% | 6% | 4% | | |
| Average | 84% | 84% | | | 82% | 89% | 80% | | |
| Below Average | 5% | 7% | | | 0% | 5% | 16% | | |

* $p < .05$; ** $p < .01$; *** $p < .001$

There was also a significant I.Q. difference-- the above average students ($\bar{X}$=32.19) scored higher than the average students ($\bar{X}$= 26.96) who scored higher than the below average students ($\bar{X}$=24.45).

Spelling

There was no significant difference between the Immersion students ($\bar{X}$=23.03) and the English Control students ($\bar{X}$=22.70). In terms of percentile, the average group score was at the 56% level.

There was a significant I.Q. difference-- again, the above average students ($\bar{X}$=25.36) scored higher than the average students ($\bar{X}$= 23.17) who scored higher than the below average students ($\bar{X}$=20.06).

English Writing Skills

Separate two way analyses of variance (unweighted means solution for unequal n) were performed on the ratings for the following scales: spelling, sentence accuracy, sentence complexity & variety, organization, originality, and overall. The main variables were School Group (Immersion, English Control) and I.Q. (above average, average, below average). The ratings from the remaining scales--punctuation, length, and vocabulary-- were not analyzed statistically because there was an insufficient number of response categories to justify use of analysis of variance. Instead, percentage frequency counts were calculated for each response category of each scale; namely "appropriate" and "inappropriate" for punctuation and length, and "superior", "average" and "below average" for vocabulary. The percentage frequency counts represent the percentage of total possible ratings that were of a given category. A summary of all analyses is presented in Table 3.

There were no statistically significant differences between the Immersion and English Control group on any of the six scales that were

analyzed statistically. Thus, the compositions of the Immersion students were regarded as comparable to those of the English Control students on these six dimensions.

There were statistically significant differences among the I.Q. groups on all of the scales, except spelling. In those five cases where there were differences, the rank order of the groups from highest rating to lowest was always above average, average and below average.

In general, the majority of the ratings on all scales were in the 2 and 3 range which signifies respectively, "below grade 5 level" to "meets expectations of grade 5 level".

The immersion and English Control students were rated comparable on the dimensions of length and vocabulary: on length, the Immersion group received 69% appropriate ratings and the English Control group received 65% "appropriate" ratings; on vocabulary, the Immersion group received 11% "superior" ratings and 84% "average", and the English Control group received 9% "superior" and 84% "average" ratings. On punctuation, the English Control group received more "appropriate" ratings (84%) than did the Immersion group (77%).

In last year's grade 4 writing evaluation, we found that the Immersion students were rated lower than the English Controls on spelling but higher on originality. These differences were not found this year in grade 5. As well, last year the Immersion students received more "appropriate" ratings than did the English Control students on punctuation, length, and vocabulary (to a lesser extent). The grade 4 differences between the two groups on length and vocabulary had disappeared this year, and the punctuation results were reversed, with the English Control students being rated higher than the Immersion students this year.

Summary

There was no evidence that the Immersion students were behind the English Control students on any of the measures of English language competence employed in this investigation-- reading, vocabulary, spelling and writing. The spelling problem which has characterized the Immersion group in previous evaluations was not evident in this evaluation.

The finding that the Immersion students scored significantly higher than the English Control students on the reading test might be interpreted to reflect the relatively smaller below average I.Q. sample in the Immersion group (n= 7 for Immersion below average, and n= 20 for the English Control below average). In other words, when the average reading scores for the school groups are calculated, the scores from all I.Q. levels are totalled and since there are relatively fewer low I.Q. students in the Immersion group, than in the English Control group, this will bias the calculation in their favor. However, when the reading data are broken down by school and I.Q. group (see Table 4 below), it is in fact the low I.Q. students from the two groups who showed the largest differential, with the Immersion students scoring higher.

| | | IMMERSION | ENGLISH CONTROL |
|---|---|---|---|
| I.Q. | Above Average | 33.22 | 31.17 |
| | Average | 26.87 | 27.05 |
| | Below Average | 27.00 | 21.90 |

TABLE 4

AVERAGE READING SCORES

Furthermore, the Immersion above average group (N= 18) also scored higher than the English Control above average group (N= 18). Thus, an explanation of the difference in reading scores in terms of sample size does not seem to be adequate. A similar difference in reading scores was found in last year's evaluation of a grade 7 Immersion group (see Genesee, Morin & Allister, 1974). One alternative explanation of these results is that instruction in a second language enhances reading comprehension in the native language, for some obscure reason. A more explicit explanation than this would be mere conjecture.

The results from this year's writing evaluation are consistent with last year's insofar as they indicate that the Immersion students are not falling behind in their English writing skills. Again, as mentioned above, a spelling problem was not evident in this year's results. The superiority that the Immersion students had demonstrated last year in length, punctuation and vocabulary skills was not evident this year.

## French Language Tests

Two separate sets of statistical analyses were performed on the results from the French language tests. In Analysis 1, the French Control group was included along with the Immersion group and the English Control group. However, since there was no I.Q. information on the French Control students, we had to perform these analyses ignoring I.Q. distinctions. The results of these analyses give us an indication of the performance of the Immersion and the English Control students relative to one another and relative to the French students. In Analysis 2, the French Control results were omitted. This allowed us to distinguish among the above average, average and below average students from the Immersion and English Control samples, thereby giving us an indication of the influence of I.Q. level on second language learning.

### ANALYSIS 1

Separate analyses of variance were performed on the scores from each of the following tests: Test de Lecture (vocabulaire), Test de Lecture (compréhension), Test de Rendement en Français (V), Test de Rendement en Mathématiques (V), compréhension aurale, and production orale: compréhension aurale, prononciation, grammaire, vocabulaire and communication. The only variable in these analyses was School Group (Immersion, English Control, French Control). A summary of these results is presented in Table 5A. The results from each test will now be discussed.

### Test de Lecture

There was a significant difference among the school groups on the vocabulaire subtest-- the French Control students ($\bar{X}$= 24.90) scored highest, followed in descending order by the Immersion students ($\bar{X}$= 20.21) and the English Control students ($\bar{X}$= 12.09).

TABLE 5A

ANALYSIS 1

Summary of French Language Test Results

| | (A) By School Group | | | | |
|---|---|---|---|---|---|
| | Immersion | English Control | French Control | F Ratio | df |
| Test de Lecture: | | | | | |
| Vocabulaire | 20.21 | 12.09 | 24.90 | 214.79*** | 2,156 |
| Compréhension | 21.33 | 13.92 | 22.47 | 97.68*** | 2,156 |
| Test de Rendement en Français | 3.47 | 1.63 | 6.7 | 67.11*** | 1,85[1.] |
| Test de Rendement en Mathématiques | 5.36 | 3.04 | 6.4 | 57.82*** | 1,88[1.] |
| Compréhension aurale | 14.62 | 7.21 | 15.87 | 241.75*** | 2,155 |
| Production Orale: | | | | | |
| Compréhension aurale | 4.92 | 3.67 | 5.00 | 28.64*** | 2,91 |
| Prononciation | 4.00 | 2.01 | 5.00 | 87.60*** | 2,91 |
| Grammaire | 4.00 | 2.02 | 5.00 | 76.88*** | 2,91 |
| Vocabulaire | 3.59 | 1.74 | 5.00 | 128.99*** | 2,91 |
| Communication | 3.65 | 2.53 | 5.00 | 212.26*** | 2,91 |

1. These analyses did not include the results from the French Control group.

*** $p < .001$

There was also a significant difference among the school groups on the compréhension subtest-- in this case, the English Control group ($\bar{X}$= 13.92) scored lower than both the French Control group ($\bar{X}$= 22.47) and the Immersion group ($\bar{X}$= 21.33). There was no difference between the latter two groups.

Test de Rendement

There was a significant difference between the Immersion and English Control groups on the Test de Rendement en Français -- the Immersion group ($\bar{X}$= 3.47) scored higher than the English Control group ($\bar{X}$= 1.63); both scored lower than the French Control group ($\bar{X}$= 6.7). These scores are presented in stanine form; an interpretation of a stanine scale is presented as Appendix D.

There was also a significant difference between the Immersion and English Control groups on the Test de Rendement en Mathématiques -- the Immersion group ($\bar{X}$= 5.36) scored higher than the English Control group ($\bar{X}$= 3.04); both scored lower than the French Control group ($\bar{X}$= 6.4).

Compréhension Aurale

There was a significant difference among the school groups on this test-- the English Control group ($\bar{X}$= 7.21) scored significantly lower than both the French Control group ($\bar{X}$= 15.87) and the Immersion group ($\bar{X}$= 14.62). Considering that the maximum possible score on this test is 17, the performance of the Immersion students is very good.

Production orale

There were significant differences among the school groups on all 5 scales of this evaluation. The general pattern of ratings from highest to lowest was always French Control, Immersion and English Control. The

ratings of the Immersion students on compréhension aurale ($\bar{X}$= 4.92), prononciation ($\bar{X}$= 4.00) and grammaire ($\bar{X}$= 4.00) are particularly noteworthy (a rating of 4 indicates excellent performance for a non-native speaker).

## ANALYSIS 2

This time, separate analyses of variance were performed on the scores from the French language tests of the Immersion and English Control groups only. The main variables in these analyses were School Group and I.Q. (above average, average, below average). Only the results pertaining to I.Q. will be reported here since the results pertaining to the school groups are essentially the same as those from Analysis 1. A summary of Analysis 2 is presented in Table 5B.

There were significant differences among the I.Q. groups on both subtests of the Test de Lecture (vocabulaire and compréhension), and on both versions of the Test de Rendement (en Français and en Mathématiques). In every case, the above average group obtained the highest score and the below average group obtained the lowest score. There was also a significant I.Q. difference on the vocabulaire scale of the oral production evaluation-- the below average group scored lower than both the average and the above average groups, which were rated equivalent.

There were no significant differences among the I.Q. groups on the remaining tests: group compréhension aurale, individual compréhension aurale, prononciation, grammaire and communication. A similar pattern of results was found last year when performance on the tests of academic language skills (reading, Test de Rendement en Français and en Mathématiques) varied as a function of I.Q. level, while the tests of interpersonal communication skills (compréhension aurale, production orale) did not.

TABLE 5B

ANALYSIS 2

Summary of French Language Test Results

(B) By I.Q.

| | Above Average | Average | Below Average | F Ratio | df |
|---|---|---|---|---|---|
| Test de Lecture: | | | | | |
| Vocabulaire | 17.66 | 15.90 | 14.72 | 5.19** | 2,96 |
| Compréhension | 19.40 | 16.95 | 16.44 | 6.14** | 2,96 |
| Test de Rendement en Français | 3.16 | 2.54 | 1.96 | 9.56*** | 2,85 |
| Test de Rendement en Mathématiques | 5.15 | 4.03 | 3.41 | 11.02*** | 2,85 |
| Compréhension Aurale | 11.29 | 10.71 | 10.56 | .85 | 2,97 |
| Production orale | | | | | |
| Compréhension | 4.30 | 4.51 | 3.99 | 1.39 | 2,57 |
| Prononciation | 3.20 | 3.19 | 2.64 | 1.66 | 2,57 |
| Grammaire | 3.30 | 3.17 | 2.56 | 2.25 | 2,57 |
| Vocabulaire | 2.87 | 2.87 | 2.15 | 3.50* | 2,57 |
| Communication | 3.14 | 3.20 | 2.88 | 1.69 | 2,57 |

* $p < .05$; ** $p < .01$; *** $p < .001$

Summary

The results indicate that the French Immersion students who have reached the grade 5 level continue to surpass the English Control students in French language competency, but at the same time do not achieve native-like mastery. Their achievements are particularly noteworthy in the areas related to interpersonal communication-- listening and speaking skills. The Immersion students are generally conversant and creative speakers of French. Their command of academic French language skills is also commendable-- they scored as well as the French Controls on reading comprehension, and they scored in the average stanine on the French math test and in the below average stanine on the Test de Rendement en Français. Their performance on the Test de Rendement en Français has fallen from stanine 4 last year to stanine 3 this year. This decline may reflect differences in testing time-- last year's test was administered in June, while this year's test was administered in November to coincide with the C.E.C.M. testing. In conclusion, this year's results are consistent with last year's.

## CONCLUSIONS AND SUMMARY

As in previous evaluations there is no evidence that the Immersion students have suffered any set backs to their English language skills; their reading, spelling, vocabulary and writing skills are comparable to those of students following the regular English program with French-as-a-second language. The differences reported last year between the Immersion and English Control groups on the writing evaluation were not as evident this year. There was a difference this year between the Immersion students and English Control students on the reading subtest, in favor of the Immersion group; this difference was not found last year.

As in previous evaluations, the Immersion students were superior to the students in the regular program in all aspects of the French language that were evaluated-- reading, language usage, listening comprehension, oral production, and mathematical terminology.

The Immersion students performed favorably when compared to the French Control students on reading comprehension, listening comprehension, some aspects of oral production (pronunciation and grammar) and mathematics.

As in last year's evaluation, we found that performance on the tests of academic French language skills (Test de Lecture, Test de Rendement en Français and en Mathématiques) depended upon the I.Q. level of the students such that the above average students scored higher than the average students, and so on. No similar relationship was found between I.Q. level and performance on the tests of interpersonal communication skills (listening and speaking). These results suggest that acquisition of those aspects of a second language which are not explicitly taught is not necessarily dependent upon the learner's intellectual level. These results suggest further that in a second language

program where the goals pertain to interpersonal communication skills, a non-academic approach to language teaching, such as characterizes the French immersion program, will be more effective with students possessing a broad range of intellectual and therefore scholastic abilities than programs which stress an academic approach.

Finally, there was no evidence that the students in the Immersion program had developed psychological problems related to feelings of Security (see Appendix E). In fact, according to their performance on the Security Test the Immersion students scored higher on security than did the non-immersion students, and were less likely to use immature responses in situations which required a decision.

An additional evaluation of this group of Immersion students will be undertaken in 1975-76, their final year in the elementary school French immersion program.

## REFERENCES

Barik, H.C., Swain, M., & McTavish, K. Evaluation of the 1972-73 French immersion Kindergarten and grade one classes at Allenby Public School in Toronto. Bilingual Education Project, mimeo, Modern Language Center, O.I.S.E., 1973.

Edwards, H.P., & Casserly, M.C. Evaluation of second language programs in the English schools- Annual report 1972-73. Ottawa Roman Catholic Separate School Board, 1973.

Genesee, F., Morin, S., & Allister, T. Evaluation of the 1973-74 Grade IV French-Immersion Class. A report submitted by the Curriculum Department to The Protestant School Board of Greater Montreal, September, 1974.

Grapko, M.F. Institute of Child Study Security Test (Elementary Form). Toronto: University of Toronto, 1957.

Lambert, W.E., & Tucker, G.R. The Bilingual Education of Children: The St. Lambert Experiment, Rowley, Mass.: Newbury House, 1972.

Lorge-Thorndike Intelligence Tests (Canadian Version), Toronto: Thomas Nelson & Sons, 1967.

Polich, E. Report on the Evaluation of the Lower Elementary French Immersion Programme through Grade III, 1971-1973. A report submitted to The Protestant School Board of Greater Montreal, December 1973.

## APPENDIX A

### SUMMARY OF 1973-74 ENGLISH-LANGUAGE TEST RESULTS

| TEST | SCHOOL GROUP | | | | I.Q. GROUP | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Immersion | Non-Immersion | F | P | Above Average | Average | Below Average | F | P |
| Reading | 77.97 | 77.30 | .13 | .72 | 84.60 | 76.67 | 71.64 | 16.47 | <.001 |
| Word Knowledge | 77.02 | 76.90 | .004 | .96 | 85.13 | 74.60 | 71.15 | 16.24 | <.001 |
| Arithmetic Problem Solving and Concepts | 84.24 | 82.47 | 1.31 | .26 | 90.49 | 83.96 | 75.63 | 30.77 | <.001 |
| Writing Skills: | | | | | | | | | |
| Spelling | 3.00 | 3.46 | 7.80 | <.05 | 3.59 | 3.18 | 2.92 | 5.57 | <.01 |
| Sentence Accuracy | 2.61 | 2.42 | 2.35 | >.05 | 3.03 | 2.37 | 2.15 | 9.92 | <.01 |
| Sentence Complexity | 2.53 | 2.36 | 2.05 | >.05 | 2.95 | 2.36 | 2.01 | 12.18 | <01 |
| Organization | 2.66 | 2.43 | 2.17 | >.05 | 3.01 | 2.43 | 2.20 | 8.50 | <.05 |
| Originality | 2.20 | 2.01 | 4.77 | <.05 | 2.36 | 2.02 | 1.95 | 5.33 | <.05 |
| Overall | 2.57 | 2.37 | 2.96 | >.05 | 2.99 | 2.37 | 2.00 | 10.51 | <.01 |
| Punctuation: | | | | | | | | | |
| Appropriate | 83% | 61% | | | | | | | |
| Inappropriate | 17% | 39% | | | | | | | |
| Length: | | | | | | | | | |
| Appropriate | 70% | 49% | | | | | | | |
| Inappropriate | 30% | 51% | | | | | | | |
| Vocabulary: | | | | | | | | | |
| Superior | 11% | 7% | | | | | | | |
| Average | 84% | 80% | | | | | | | |
| Below Average | 5% | 13% | | | | | | | |

## APPENDIX B

### SUMMARY OF 1973-74 FRENCH-LANGUAGE TEST RESULTS (INCLUDING I.Q. RESULTS)

| TEST | SCHOOL GROUP | | | | | I.Q. GROUP | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Immersion | Non-Imm | French | F | P | Above | Average | Below | F | P |
| I.Q. | 108.67 | 105.68 | 106.36 | 2.70 | .07 | 124.81 | 106.59 | 89.31 | 346.33 | <.001 |
| Reading | | | | | | | | | | |
| Test 1 | 38.17 | 30.83 | 42.55 | 69.07 | <.001 | 39.72 | 36.63 | 35.21 | 10.47 | <.001 |
| Test 2 | 14.78 | 9.54 | 15.54 | 38.06 | <.001 | 15.42 | 12.68 | 11.76 | 17.93 | <.001 |
| T. de R. en Français | | | | | | | | | | |
| Raw scores | 25.10 | 11.17 | 32.52 | 244.69 | <.001 | 25.42 | 23.61 | 19.76 | 17.44 | <.001 |
| Stanines | 4.36 | 1.33 | 6.56 | 208.00 | <.001 | 4.60 | 4.28 | 3.36 | 12.47 | <.001 |
| T. de R. en Math. | | | | | | | | | | |
| Raw scores | 23.52 | 12.06 | 23.22 | 49.40 | <.001 | 23.07 | 19.59 | 16.15 | 13.84 | <.001 |
| Stanines | 5.20 | 2.19 | 4.95 | 47.28 | <.001 | 5.09 | 4.04 | 3.21 | 14.99 | <.001 |
| Listening Comprehension | 30.30 | 16.61 | 30.94 | 301.48 | <.001 | 26.95 | 26.15 | 24.74 | 5.75 | .003 |
| Oral Production | | | | | | | | | | |
| Comprehension | 4.97 | 3.32 | 5.00 | 47.07 | <.01 | 4.49 | 4.52 | 4.28 | .72 | > .05 |
| Pronunciation | 4.03 | 3.17 | 4.92 | 51.51 | <.01 | 3.92 | 4.28 | 3.93 | 5.28 | < .05 |
| Grammar | 4.49 | 2.64 | 4.91 | 50.46 | <.01 | 4.14 | 4.13 | 3.77 | 1.49 | > .05 |
| Vocabulary | 4.81 | 2.36 | 5.00 | 92.43 | <.01 | 4.10 | 4.24 | 3.83 | 1.23 | > .05 |
| Communication | 4.61 | 3.12 | 4.94 | 79.40 | <.01 | 4.32 | 4.29 | 4.06 | 1.96 | > .05 |

## APPENDIX C

### EXPLANATION OF STATISTICAL TERMINOLOGY

Analysis of variance is a mathematical procedure for testing the reliability of the differences among a series of average scores. Stated simply, an analysis of variance estimates how much of the variability in the test scores is due the systematic effects of the experimental treatment which in the current report is the school language program and how much of the variability is due to non-systematic influences such as individual differences among the children. The first estimate is generally called the "treatment variance" and the second estimate is called the "error variance". The ratio of the treatment variance to the error variance is called an F ratio. If the treatment variance is larger than the error variance then the F ratio will be greater than 1. If the error variance is larger than the treatment variance then the F ratio will be less than 1. A very small F ratio indicates that most of the variability in the scores is due to unsystematic individual differences which are unrelated to the particular treatment in question..

To decide whether a particular F ratio is"significant" and, in other words whether there is a significant difference between the average scores, one must determine the probability (or p value) of obtaining an F ratio of a certain magnitude. The p value can be ascertained easily with the assistance of standard mathematical tables, that have already been set up, and a knowledge of the"degrees of freedom" for your particular experiment. There is a degree of freedom value associated with the numerator, or treatment variance, of the F ratio, and a degree of freedom value associated with the denominator, or error variance, of the F ratio. There is no brief way of describing the calculation or significance of degrees of freedom

so that for the sake of this explanation they will not be described further.

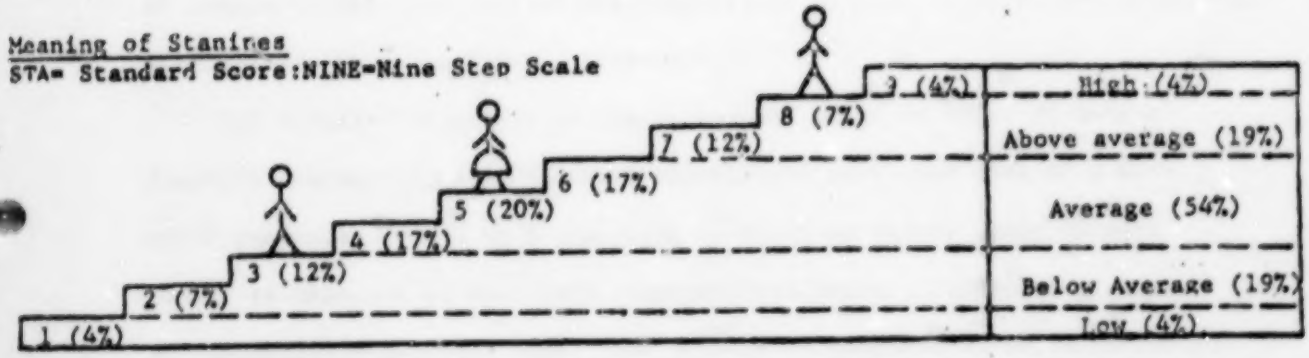
According to tradition, the probality (or p value) associated with a particular F ratio must be 5 out of 100 (i.e., .05) or less ($p \leq .05$) before one calls the difference between the averages "significant". If the p value is greater than 5 out 100 (i.e. $p > .05$), then one calls the difference between the averages "non-significant". What one is saying when one says "$p < .05$" is that the probability of getting a particular F ratio, and, therefore, a particular difference between average scores is less than 5 in 100 by chance alone . One then goes on to assume that since the probability is relatively low, that the difference between the mean scores is probably due to the treatment and not to chance or uncontrolled influences. In other words, the treatment (or language of schooling in this report) has had a significant influence on the students' performance on the language test. Of course, the converse rationale holds for the reverse situation where $p > .05$.

This explanation of analysis of variance, F ratios, p values and significance is of necessity brief, but hopefully will facilitate the reader's understanding of the report. More thorough and mathematical descriptions of these concepts can be found in Statistical Analysis for Psychology and Education, 3rd ed.,G.A. Ferguson, McGraw-Hill, 1971, and Statistical Principles in Experimental Design, 2nd ed., B.J. Winer, McGraw-Hill, 1971.

## INTERPRETATION OF STANINE SCORES

**Stanine** - The stanine scale is a simple nine point scale of standard scores. (The word stanine was derived for STAndard NINE-point scale). In this scale, raw scores are converted to scores which range from 1 (low) to 9 (high) with a mean of 5 and a standard deviation of 2.

**Figure 1. Percentage of Cases at Each Stanine Level**



## APPENDIX E

### Security Test

This is a group test designed to assess security. Security is defined by the author of this test (Grapko, 1957) as "the child's willingness to accept consequences for his decision or behavior" (pg. 1). The test is presented in the form of a story which describes the events in the day of the life of a little boy named Jimmy. " These events include 1) his relationships to family, teachers and friends, 2) his attitudes and practices relating to school work, and his handling of money, 3)his avocational interests and play, and 4) his response to physical injury, and orientation with respect to an evolving philosophy."

At a number of points in the story, the child is asked to make a decision taking the role of Jimmy; the student must rank each of 5 possible responses from 1 to 5 according to which he thinks Jimmy is most likely to make,and so on. Each response represents a different concept:

1) Independent Security (IS): the ability to complete an activity and the willingness to accept one's own decisions, actions and consequences in the performance of the activity;

2) Mature Dependent Security (MDS): the willingness to share in the performance of an activity combined with a willingness to mutually accept the consequences of decisions and actions;

3) Immature Dependent Security (IDS): to wait for or expect help in completing a task or performing an activity;

4) Deputy Agent (DA): the avoidance of consequences by means of some "psychological shuffle"; e.g. blaming someone else, making up a false excuse, etc;

5) Insecurity (INS): to lack skill in dealing with an activity or significant "event" which gives rise to indecision, hesitation and anxiety.

From the student's responses to the decision points, a consistency score is calculated which measures the degree of concordance or uniformity that the child shows in ranking the decision statements. The higher the score, the higher the consistency; the consistency score ranges from 0 to 100.

A security score is also calculated based on the 5 response categories above. The security score is a measure of the degree to which the child's rank order of decisions agrees with the "ideal" order of rank as determined by security theory. Thus, the child who ranks the independent secure items first, the mature dependent decisions second, the immature dependent secure decisions third, the deputy agent decisions fourth, and the insecure decisions fifth is in accord with the "ideal" order of rank. The security scores range from a minimum of 1 to a maximum of 100.

## Security Test Results

Separate analyses of variance were performed on the main scores of consistency and security, and on the sub-scores of Independent Security, Mature Dependent Security, Immature Dependent Security, Deputy Agent and Insecurity. None of the scores recorded in Table 5 resemble the actual priority values assigned in the test, since they are the result of mathematical transformations of a relatively complex nature. In the case of the consistency and security scores, a high score indicates high consistency or high security. In the case of the remaining 5 scores, the higher the value, the lower the rank priority assigned to the response category. For example, in Table 6, which presents a summary of the security test results by school group, sex and I.Q., a consistency score of 30.85 (Immersion group) indicates higher response consistency than a score of 23.29 (English Control). On the other hand, an insecurity value of 53.29 for the Immersion Group indicates that these students generally gave responses in this category a lower rank priority than did the English Control students with an average insecurity score of 44.35. This suggests that the Immersion students were less likely to make decisions which might be considered indicative of insecurity than did the English Control students.

Three sets of analyses were carried out; first of all looking at differences between the school groups; secondly looking at differences between boys and girls; and finally looking at differences among the I.Q. groups (see Tables 6A, B and C respectively).

The main interest in these tables is with the CONSISTENCY and SECURITY scores. Inspection of the values in each sub-category, gives an indication of specific response styles which contribute to the overall consistency and security scores.

(A) School Group

There was a significant difference between the Immersion and English Control group on both the consistency and the security scores (The author of the test points out that the correlation between these two scores tends to be high -- r =·87. In both cases the Immersion students (Consistency= 30.85; Security= 66.05) received higher scores than the English Control students (Consistency= 23.29; Security= 54.48). These results suggest two general conclusions: 1) the Immersion students have a stronger and better organized decision system than do the English Control students; and 2) the Immersion students are more likely than the English Control students to make decisions reflecting feelings of security.

Further examination of the sub-category values indicates that the difference between the two groups may be due to the tendency of the Immersion students to give lower priorities to the decisions reflecting immature dependent security, deputy agent responses and insecurity than did the English Control students. The priorities accorded to the decisions reflecting independent security and mature dependent security were fairly comparable between the groups.

Post hoc testing of this sort is subject to two interpretations:

1) that the obtained differences existed before the students were exposed to their respective programs; and

2) that the obtained differences reflect consequences of exposure to the different programs. In either case, there is no evidence from this particular test which assesses a particular psychological concept that the Immersion students have suffered ill effects of a psychological sort from the French immersion experience.

TABLE 6

SUMMARY OF SECURITY TEST RESULTS

(A) by School Group

| | Immersion | English Control | F Ratio [1.] |
|---|---|---|---|
| CONSISTENCY | 30.85 | 23.29 | 3.68* |
| SECURITY | 66.05 | 54.48 | 4.25* |
| 1) Independent Security | 32.58 | 27.86 | 2.45 |
| 2) Mature Dependent Security | 32.42 | 28.87 | 1.36 |
| 3) Immature Dependent Security | 45.82 | 37.13 | 6.36* |
| 4) Deputy Agent | 54.08 | 44.56 | 5.52* |
| 5) Insecurity | 53.29 | 44.35 | 4.92* |

1. df = 1, 103

* $p < .05$

(B) Sex

As was pointed out earlier, there were too few males and females in the below average Immersion group to look at the School Group by Sex interactions. Consequently, when we analyzed sex differences on the Security test, we ignored distinctions between school groups. There were no significant differences between the boys and the girls on either of the main scores or on any of the sub-scores. This would suggest that the boys and girls, regardless of language program, were developing security feelings in the same ways and at the same rates.

(C) I.Q.

There were significant differences among the I.Q. groups on the consistency score-- the below average group ($\bar{X}$= 22.11) had a lower consistency score than either the average ($\bar{X}$= 33.28) or above average group ($\bar{X}$= 32.97). This suggests that the below average students as a group had a weaker and less well organized decision system than did the other students. There is further evidence from the results on the sub-categories that the below average group's decision system was less well organized: the below average group ($\bar{X}$= 39.46) gave mature dependent security responses relatively high priority when compared to the average ($\bar{X}$= 34.13) and above average students ($\bar{X}$= 32.44), and they also gave deputy agent ($\bar{X}$= 51.89) and insecurity responses ($\bar{X}$= 51.18) relatively high priority relative to the average group (deputy agent= 55.81; insecurity= 55.64) and the above average group (deputy agent= 56.50; insecurity= 55.43). These two types of decisions are conceptually incompatible so that co-occurrence of these two kinds of decisions in a child's repertoire could lead to less mature decision making. Evidently there is no difference between the below average students and the other students on the security score because the relatively high priority that these two types of responses received by the English Control students cancelled one another out.

TABLE 6

SUMMARY OF SECURITY TEST RESULTS

| | (B) by Sex | | | (C) by I.Q. | | | |
|---|---|---|---|---|---|---|---|
| | Male | Female | F Ratio [1] | Above Average | Average | Below Average | F Ratio [2] |
| CONSISTENCY | 28.10 | 26.04 | .27 | 32.97 | 33.28 | 22.11 | 4.04* |
| SECURITY | 61.77 | 58.76 | .29 | 68.61 | 67.95 | 63.11 | .78 |
| 1) Independent Security | 28.81 | 31.63 | .87 | 33.65 | 33.05 | 36.11 | 1.06 |
| 2) Mature Dependent Security | 29.74 | 31.54 | .35 | 32.44 | 34.13 | 39.46 | 6.45** |
| 3) Immature Dependent Security | 41.20 | 41.75 | .03 | 46.98 | 46.38 | 44.21 | 1.90 |
| 4) Deputy Agent | 49.66 | 48.98 | .03 | 56.50 | 55.81 | 51.89 | 3.19* |
| 5) Insecurity | 48.31 | 49.32 | .06 | 55.43 | 55.64 | 51.18 | 3.21* |

1. df= 1, 103
2. df= 1,90

* $p < .05$ ** $p < .01$